

A personalized, integrated system for rice functional genomics

NETTAB (Network Tools and Applications in Biology) 2007, 2007/06/12-15, Pisa, Italy

S. FROMENTIN, G. DROC and P. LARMANDE
CIRAD, Biological Systems Department, UMR 1098 DAP,
Avenue Agropolis, F-34398 Montpellier 05, France
Corresponding author : pierre.larmande@cirad.fr

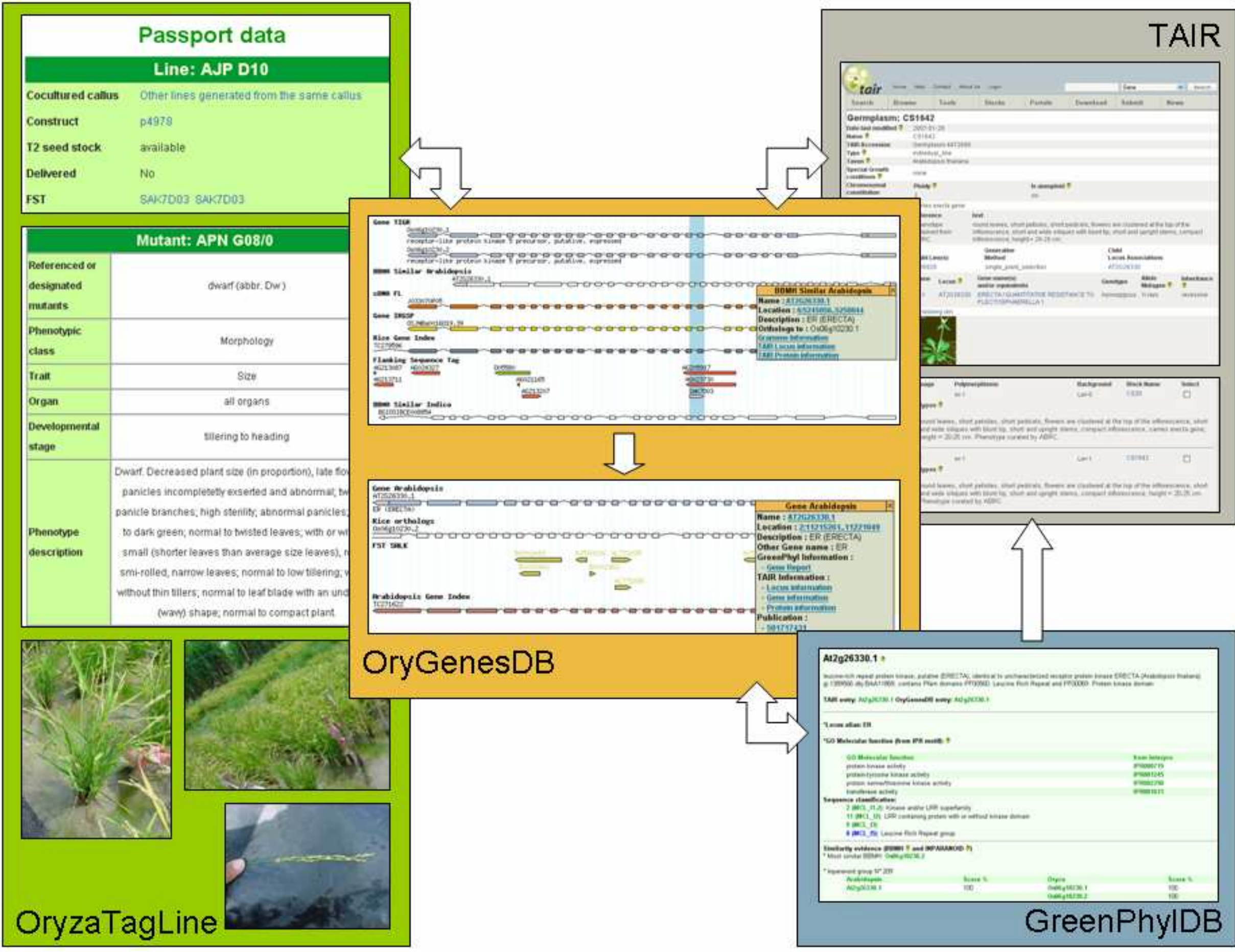
Introduction

We recently developed three plant databases specifically designed for assisting researches in rice functional genomics :

- OryGenesDB [1], a reverse genetics and genomic database ,
- OryzaTagLine [2], a phenotypic database describing mutant insertion lines,
- GreenPhylDB [3], a database for comparative functional genomics to link the two models plant species rice and Arabidopsis through ortholog predictions.

Scientists need flexible systems which reproduce automatically their actions and generate synthetic results.

The aim of this project is to design and implement an architecture which retrieve data from these three remote databases, store queries and results, and execute users queries periodically.



Number	Web Service Name	Source	Input	Output
1	getInfoByGeneId	OryGenesDB	Gene id(s)	Full record
2	getFSTbyGeneId	OryGenesDB	Gene id(s)	FST id(s)
3	getGeneIdByLocation	OryGenesDB	Location	Gene id(s)
4	getGeneIdByFST	OryGenesDB	FST id(s)	Gene id(s)
5	getPhenotypeByPlant	OryzaTagLine	Plant name	Full record
6	getExpressionByPlant	OryzaTagLine	Plant name	Full record
7	getFSTByPlant	OryzaTagLine	Plant name	FST id(s)
8	getGeneIdByInterpro	GreenPhylDB	Interpro domain	Gene id(s)
9	getInfoByGeneId	GreenPhylDB	Gene id(s)	Full record
10	getPlantByFST	OryzaTagLine	FST id	Plant name

Service definition

We defined several Web Services according to data that are required to :

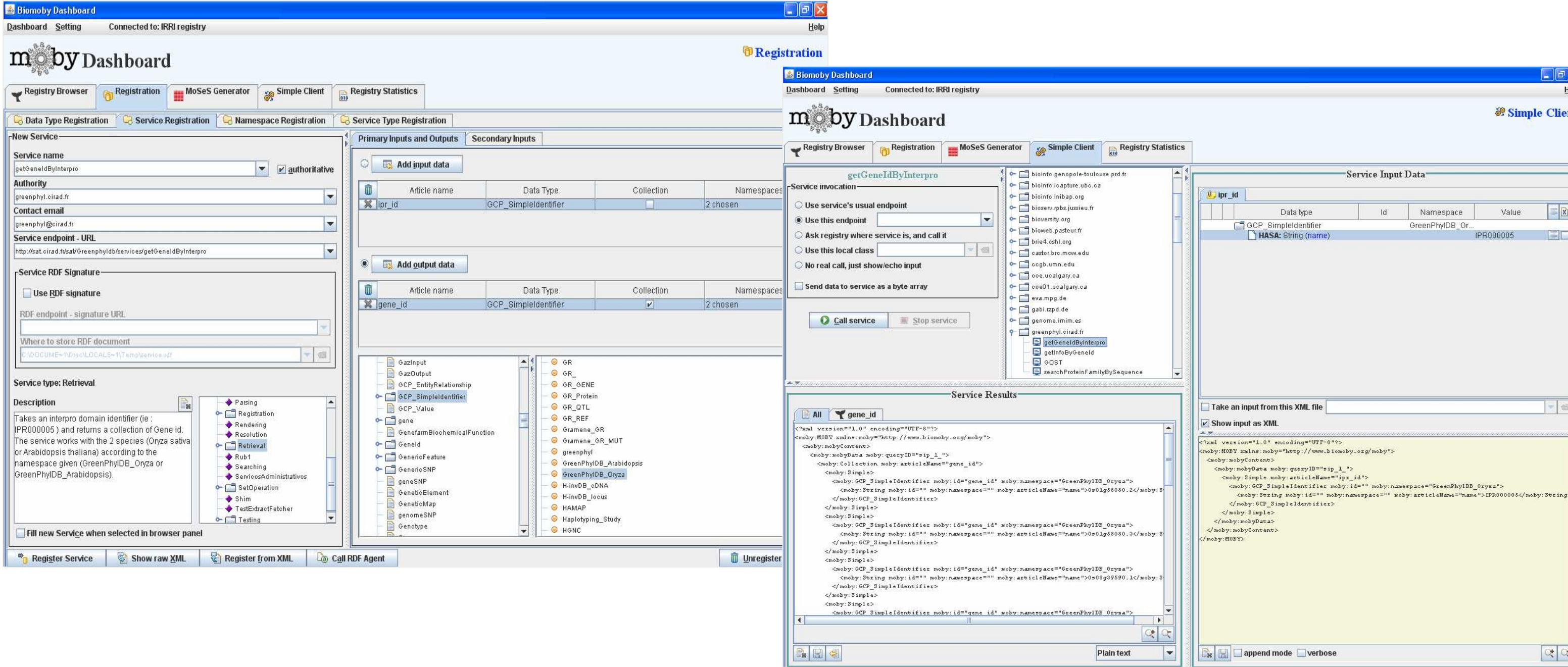
- extract information from these databases,
- create interconnections between them.

The table lists all Web Services that have been created from our databases. Gene id refers to the Locus Name defined by TIGR and/or TAIR database. Location refers to genomic coordinates. FST id refers to the Flanking Sequence Tag mapped on rice genome and is directly linked to a plant name. Plant name refers to OryzaTagLine.

Service deployment

Web Services were generated using BioMoby dashboard [4]. First, we registered namespace and services for each database, and we used data types defined by the Generation Challenge Programme [5], like GCP_SimpleIdentifier or GCP_Feature.

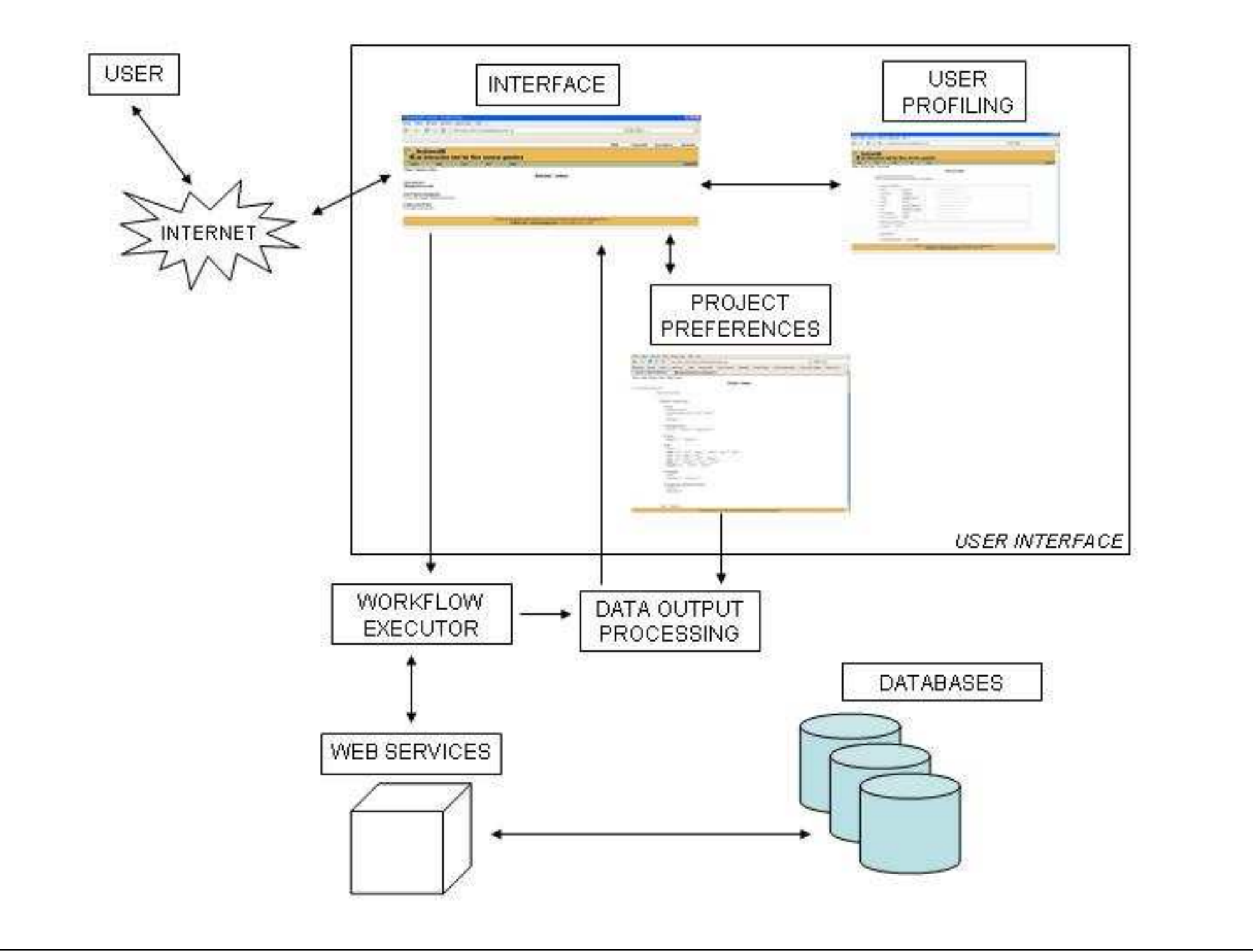
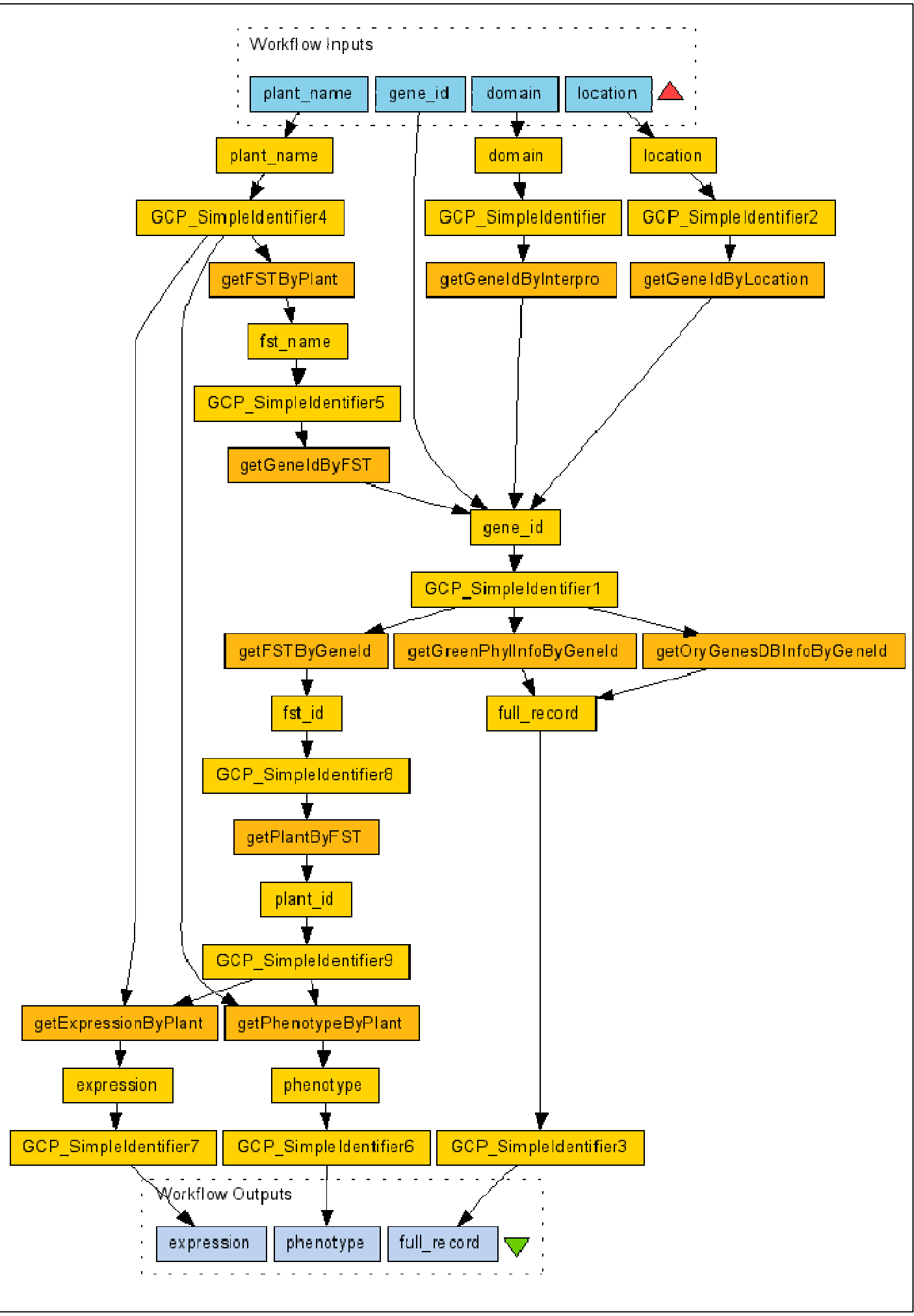
Next, with the Moses module, we generated data types and skeletons, that we used to implement with Eclipse. Finally, the “Single Client” panel allows us to call our services and display results.



Results

We developed a web interface to create user accounts and register user profiles. All profiles and preferences are stored in a MySQL data management system and queries are automatically and periodically executed.

With the BioMoby plug-in of Taverna [6], we designed a set of predefined workflows, thus simplifying access to end users. Users can initiate the workflow according to 4 inputs : **location**, **interpro domain**, **gene id** and **plant name**. Entries like "gene_id" and "fst_name" make it possible to link our databases. Workflows have been designed to access and retrieve pertinent information from the different databases (phenotype description, expression, rice Arabidopsis gene ortholog information...).



Conclusion

Using this architecture, biologists can automate their web queries therefore reducing the time wasted on several web interfaces to extract the pertinent information. With the stored preferences, users can personalize the system to display synthetic relevant information. They can also be alerted when updates are available on the different databases.

References :

- [1] <http://orygenesdb.cirad.fr>
- [2] <http://urgi.versailles.inra.fr/OryzaTagLine/>
- [3] <http://greenphyl.cirad.fr>
- [4] http://biomoby.open-bio.org/CVS_CONTENT/moby-live/Java/docs/Dashboard.html
- [5] <http://www.generationcp.org/>
- [6] <http://taverna.sourceforge.net/>

